

(19)日本国特許庁 (JP)

(12) 公表特許公報 (A)

(11)特許出願公表番号
特表2000-503774
(P2000-503774A)

(43)公表日 平成12年3月28日 (2000.3.28)

(51)Int.Cl. ⁷	識別記号	F I	テマコト [®] (参考)
G 0 1 N 21/64		G 0 1 N 21/64	F
C 1 2 N 15/09		C 1 2 Q 1/68	Z
C 1 2 Q 1/68		G 0 1 N 27/42	3 1 1
G 0 1 N 27/416		30/74	E
27/42	3 1 1	30/86	B
		審査請求 有	予備審査請求 有 (全 39 頁) 最終頁に統く

(21)出願番号 特願平10-500592	(71)出願人 ザ パーキン-エルマー コーポレーション アメリカ合衆国 カリフォルニア州 94404, フォスター シティ, リンカーン センター ドライブ 850
(86) (22)出願日 平成9年5月15日(1997.5.15)	(72)発明者 シラフ, ムハマド エイ. アメリカ合衆国 カリフォルニア 94605, オークランド, セコイヤ ビュウ ドライブ 96
(85)翻訳文提出日 平成10年12月2日(1998.12.2)	(74)代理人 弁理士 山本 秀策
(86)国際出願番号 PCT/US97/08237	
(87)国際公開番号 WO97/46963	
(87)国際公開日 平成9年12月11日(1997.12.11)	
(31)優先権主張番号 08/659, 115	
(32)優先日 平成8年6月3日(1996.6.3)	
(33)優先権主張国 米国(US)	
(81)指定国 EP(AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), AU, CA, JP	

(54)【発明の名称】 統計的信頼区間の決定を含む多成分分析方法

(57)【要約】

スペクトル的に分解可能な複数の分子種を含む混合物のスペクトル応答から得られる信号の多成分分析を行う。処理装置における信号処理方法が提供される。この方法は、各種に対する濃度推定および統計的信頼性区間の決定の両方を提供する。この方法では、多チャネル検出器からデータベクトル d が受け取られ、データベクトル d は長さ n_d を有し、 n_d はモニタされる検出器チャネルの数である。 n_d 行および n_d 列を有する較正行列 K が提供され、ここで n_d は n_d より大きく、 n_d はスペクトル的に分解可能な分子種の数である。次に、長さ n_d を有する濃度推定値ベクトル c が決定される。最後に、該濃度推定値ベクトルの該要素のそれぞれに対する信頼区間 C_{ij} が、式 $C_{ij} = C_{ij} \pm \text{varcovar}(c_{ij})^{1/2} Q_{(ij, ij)}$ に従って決定される。ここで、 Q は、所与の信頼水準 c_{ij} および所与の数の自由度 n_d に対する統計分布の臨界値であり、ここで $n_f = n_d - n_d$ である。本発明はさらに、本方法を実現するプログラム格納装置、本方法を用いるDNA配列決定プロセス、および本方法を実行する装置を含む。

【特許請求の範囲】

1. スペクトル的に分解可能な複数の分子種を含む混合物のスペクトル応答から得られる信号の多成分分析を行う、処理装置での信号処理方法であって、各種に対する濃度推定および統計的信頼区間の決定を含む方法であって、

多チャネル検出器からデータベクトル d を受け取るステップであって、データベクトル d は長さ n_c を有し、 n_c はモニタされる検出器チャネルの数であり、各ベクトル要素 d_i の大きさは特定のチャネル i での信号強度を表す、ステップと、

n_c 行および n_p 列を有する較正行列 K を提供するステップであって、 n_c は n_p より大きく、 n_p はスペクトル的に分解可能な分子種の数であり、各行列要素 K_{ij} の大きさは、特定の種 j に対する特定のチャネル i での較正值を表す、ステップと、

長さ n_p を有する濃度推定値ベクトル c を決定するステップであって、各ベクトル要素 c_i の大きさは特定の種 i の濃度推定値を表す、ステップと、

該濃度推定値ベクトルの該要素のそれぞれに対する信頼区間 C_{ii} を式

$$C_{ii} = c_i \pm (\text{varcovar}(c_{ii}))^{1/2} Q(c_i, n_f)$$

に従って決定するステップであって、ここで Q は、所与の信頼水準 c_1 および所与の数の自由度 n_f に対する統計分布の臨界値であり、ここで $n_f = n_c - n_p$ である、ステップと、

を包含する方法。

2. n_c は 10 より大きい、請求項 1 に記載の方法。

3. n_c は 20 より大きい、請求項 1 に記載の方法。

4. n_c / n_p は 1.5 より大きい、請求項 1 に記載の方法。

5. n_c / n_p は 3.0 より大きい、請求項 1 に記載の方法。

6. n_c / n_p は 5.0 より大きい、請求項 1 に記載の方法。

7. 前記統計的分布はステューデント t 分布である、請求項 1 に記載の方法。

8. 前記方法は、規定期間に規則的な区間で繰り返される、請求項 1 に記載の方法。

9. 前記濃度推定値ベクトルを決定するステップは、式

$$\mathbf{c} = (\mathbf{K}^T \mathbf{K})^{-1} \mathbf{K}^T \mathbf{d}.$$

を用いて行われる、請求項 1 に記載の方法。

10. 前記信頼区間を決定するステップは、前記濃度推定値ベクトル \mathbf{c} の残差ベクトル \mathbf{r} を決定するステップであって、ここで

$$\mathbf{r} = \mathbf{d} - \mathbf{K}\mathbf{c}$$

であるステップと、

前記データベクトル \mathbf{d} の分散 $\text{var}(\mathbf{d})$ を決定するステップであって、ここで

$$\text{var}(\mathbf{d}) = \frac{\mathbf{r}^T \mathbf{r}}{(n_c - n_p)}$$

であるステップと、

該濃度推定値ベクトル \mathbf{c} の分散—共分散行列 $\text{covar}(\mathbf{c})$ を決定するステップであって、ここで

$$\text{varcovar}(\mathbf{c}) = \text{var}(\mathbf{d}) (\mathbf{K}^T \mathbf{K})^{-1}$$

であるステップと、

を包含する、請求項 1 に記載の方法。

11. 機械によって実行可能な命令のプログラムを具体的に実現して、統計的信頼区間の決定を含む、多チャネル検出器から得られる信号の多成分分析を行う方法ステップを実行する、機械によって可読なプログラム格納装置であって、該方法ステップは、

多チャネル検出器からデータベクトル \mathbf{d} を受け取るステップであって、データベクトル \mathbf{d} は長さ n_c を有し、 n_c はモニタされる検出器チャネルの数であり、各ベクトル要素 d_i の大きさは特定のチャネル i での信号強度を表す、ステップと

n_c 行および n_p 列を有する較正行列 \mathbf{K} を提供するステップであって、 n_c は n_p より大きく、 n_p はスペクトル的に分解可能な分子種の数であり、各行列要素 K_{ij} の大きさは、特定の種 j に対する特定のチャネル i での較正值を表す、ステッ

プと、

長さ n_p を有する濃度推定値ベクトル c を決定するステップであって、各ベクトル要素 c_i の大きさは特定の種 i の濃度推定値を表す、ステップと、該濃度推定値ベクトルの該要素のそれぞれに対する信頼区間 C_{ii} を式

$$C_{ii} = c_i \pm (\text{varcovar}(c_{ii}))^{1/2} Q(c_i, n_f)$$

に従って決定するステップであって、ここで Q は所与の信頼水準 c_1 および所与の数の自由度 n_f に対する統計分布の臨界値であり、ここで $n_f = n_c - n_p$ である、ステップと、

を包含する方法。

12. ポリヌクレオチド配列決定方法であって、

(a) 第1、第2、第3および第4クラスの標識ポリヌクレオチドの混合物を、該第1クラスの各ポリヌクレオチドは3'末端ジデオキシアデノシンを含み第1の標識で標識され、該第2クラスの各ポリヌクレオチドは3'末端ジデオキシチジンを含み第2の標識で標識され、該第3クラスの各ポリヌクレオチドは3'末端ジデオキシグアノシンを含み第3の標識で標識され、そして該第4クラスの各ポリヌクレオチドは3'末端ジデオキシチミジンを含み第4の標識で標識され

るようく形成するステップであって、該第1、第2、第3および第4の標識のそれぞれはスペクトル的に分解される、ステップと、

(b) 該標識ポリヌクレオチドを電気泳動により分離して、同様な大きさのポリヌクレオチドよりなる多数のバンドを形成するステップと、

(c) n_c チャネルを有するスペクトルアレイ検出器を用いて該バンドを検出するステップであって、 n_c は4より大きい、ステップと、

(d) n_c 行および n_p 列を有する較正行列 K を提供するステップであって、 n_c は n_p より大きく、 n_p はスペクトル的に分解可能な分子種の数であり、各行列要素 K_{ij} の大きさは、特定の種 j に対する特定のチャネル i での較正值を表す、ステップと、

(e) 長さ n_p を有する濃度推定値ベクトル c を決定するステップであって、

各ベクトル要素 c_i の大きさは特定の種 i の濃度推定値を表す、ステップと、

(f) 該濃度推定値ベトルの該要素のそれぞれに対する信頼区間 C_{Ii} 式

$$C_{Ii} = c_i \pm (\text{varcovar}(c_{ii}))^{1/2} Q(c_i, n_f)$$

に従って決定するステップであって、ここで Q は所与の信頼水準 c_1 および所与の数の自由度 n_f に対する統計分布の臨界値であり、ここで $n_f = n_c - n_p$ である、ステップと、

(g) c_i および C_{Ii} の値を報告するステップと、

(i) 関連するすべてのバンドが分析されるまでステップ (c) ~ (g) を繰り返すステップと、

を包含するポリヌクレオチド配列決定方法。

13. n_p 個のスペクトル的に分解可能な分子種を含有する混合物を含むサンプルのスペクトル応答を測定する装置であって、

該サンプルのスペクトル応答を検出するスペクトルアレイ検出器であって、 n_c 個のチャネルを有し、ここで n_c は n_p より大きく、該検出器の出力は、特定の時間における特定のチャネルでの光強度を表す信号である、検出器と、

該スペクトルアレイ検出器からの信号を受け取る信号処理装置であって、以下

のステップ、

該スペクトルアレイ検出器からデータベクトル d を受け取るステップであって、該データベクトル d は長さ n_c を有し、 n_c はモニタされる検出器チャネルの数であり、各行列要素 d_{ij} の大きさは特定のチャネルでの信号強度を表す、ステップと、

n_c 行および n_p 列を有する較正行列 K を提供するステップであって、 n_c は n_p より大きく、 n_p はスペクトル的に分解可能な分子種の数であり、各行列要素 K_{ij} の大きさは、特定の種 j に対する特定のチャネル i での較正值を表す、ステップと、

長さ n_p を有する濃度推定値ベクトル c を決定するステップであって、各行列要素 c_i の大きさは特定の種 i の濃度推定値を表す、ステップと、

該濃度推定値ベクトル C_{Ii} の該要素のそれぞれに対する信頼区間を式

$$Cl_i = ci \pm (\text{varcovar}(ci))^{1/2} Q_{(ci, nf)}$$

に従って決定するステップであって、ここでQは所与の信頼水準 c_1 および所与の数の自由度 nf に対する統計分布の臨界値であり、ここで $nf = nc - np$ である、ステップと、

を行うように構成されている、信号処理装置と、
を備えた装置。

【発明の詳細な説明】

統計的信頼区間の決定を含む多成分分析方法発明の分野

本発明は、スペクトル的に区別可能な複数の種を含む混合物のスペクトル応答の同時測定において生じる信号の分析に有用な方法に関する。具体的には、本発明は、各推定値(estimate)についての統計的に有効な信頼区間の決定を含む、それぞれの種の相対的濃度の推定値の重複決定測定(overdetermined measurements)を用いる多成分分析方法に関する。

参考文献

- Caskey, C. T.、米国特許第5,364,759号 (1994)
- Connell, C. ら、*Biotechniques* 5: 342-348 (1987)
- Draper, N. および Smith, H.、"Applied Regression Analysis, 2nd Edition", John Wiley, New York (1981)
- Frans, S. D. ら、*Anal. chem.* 57: 2680-2684 (1985)
- Hunkapiller, M. ら、米国特許第4,811,218号 (1989)
- Jolliffe, I. T.、"Principle Component Analysis", p129-155, Springer-Verlag, New York (1986)
- Kalivas, J. ら、*Anal. chem.* 58: 989-992 (1986)
- Karger, A. ら、*Nucleic Acids Research* 19: 4955-4962 (1991)
- Lee, L. G. ら、*Nucleic Acids Research* 20: 2471-2483 (1992)
- Menchen, S. M. ら、米国特許第5,188,934号 (1993)
- Menchen, S. M. ら、米国特許第5,290,418号 (1994)
- Otto, M. ら、*Analytica Chimica Acta* 180: 445-456 (1986)
- Sharaf, M. A.、*Anal. chem.* 58: 3084-3091 (1986)
- Smith, L. M. ら、米国特許第5,171,534号 (1992)
- Thomas, E. V.、*Anal. chem.* 66: 795A-804A (1994)
- Weber, J. L.、米国特許第5,075,217号 (1991)

背景

しばしば、各成分種(component species)の相対濃度を決定する、スペクトル的に分解可能な複数の種を含むサンプルの分析を行うことが望まれる。このような単一サンプル混合物中の複数の種の同時検出は、それぞれ唯一の種を含む複数のサンプル混合物の連続的な分析にはない複数の利点を有する。第1に、単一のサンプル混合物のみを分析するので、サンプルの処理に必要なステップ数が比較的少なく、測定を1回だけ行えばよい。これらの特徴は共に、より高いサンプルスループットおよび向上したユーザ簡便性につながる。さらに、複数の種を組み合わせて単一の混合物にすることによって、内部較正(internal calibration)が助長される。このような同時多種スペクトル検出を用いた処理の重要な例の1つは、スペクトル的に分解可能な4つの蛍光色素(fluorescent dyes)を同時に検出する多色DNA配列決定である(Smith, Connell, Hunkapiller)。

そのスペクトル応答が少なくとも部分的には重ならないレポーターの収集物を見つけるのは困難であるので、このような同時測定全てに共通する課題は、複数の種からのスペクトル的寄与(spectral contribution)を含むデータを与えられて個々の種のそれぞれの濃度を決定することである。つまり、個々の種の濃度を決定するためには、測定されたスペクトルデータをデコンボリューションしなければならない。例えば、図1は、4色DNA配列決定において使用される4つの色素の発光スペクトルを示す。スペクトル的に純粋な信号および十分な発光強度の両方を生じる検出波長のセットを特定することが不可能であることが、これらのスペクトルから明らかである。

線形多成分分析は、スペクトル的に重なる複数の種からの寄与を含むスペクトルデータを与えられて個々の種の濃度を決定するのに有用な、強力なデコンボリューション方法である(Frans, Kalivas, Thomas)。線形多成分分析の場合、

$$d = K c + r$$

という形式の一連の一次方程式を解く。但し、 d は、その要素が、特定の波長で測定されたスペクトル応答に対応するベクトルであり、 K は、その要素が、各チャネルにおいてそれぞれの種についての純粋な成分の線形応答定数に対応する較

正行列であり、 c は、その要素が、混合物中の特定の種の濃度の推定値に対応す

る濃度推定値ベクトル(concentration estimate vector)であり、 r は、濃度推定値ベクトル c の残差ベクトル(residual vector)である。(本願の開示内容全体にわたって、行列は太字の大文字によって示され、ベクトルは太字の小文字によって示されることに留意されたい)従って、測定された値 d および較正行列 K が与えられると、それぞれの種の個々の濃度の推定値 c が決定され得る。上記方程式は、ベクトル d および r が列ベクトルであると仮定した形で書かれている。ベクトル d および r が行ベクトルとして表記される場合、先の方程式は、

$$d = K^T c^T + r$$

となる。本願の開示内容全体にわたって、ベクトル d および r は列ベクトルであると仮定する。

混合物中のそれぞれの種の濃度の推定値を求めるために加えて、その推定値の質についての定量的な効果尺度(figure of merit)を求めることが望まれる。例えば、多色DNA配列決定方法の場合、配列中の特定の塩基コール(base call)の質の尺度を有することは有用である。これは、配列中の所与の位置に2つの塩基が存在し得るヘテロ接合体位置を含む配列の場合に特に当てはまる。線形多成分分析において使用されるこのような効果尺度の最も一般的なものは、較正行列の条件数、cond(K)である。但し、

$$\text{cond}(K) = \|K\| \cdot \|K^{-1}\|$$

であり、2重括弧はその行列のノルムを示す(Otto)。

しかし、条件数は、多成分濃度推定値の質を示すための最適な効果尺度ではない。条件数は、特定の多成分測定の質の尺度ではなく、較正行列 K の質の尺度である。従って、あるシステムが有利な条件数を有し得、特定の実験ファクタに起因して結果的に特に不確定な測定となる可能性がある。例えば、条件数は、ノイズ性信号(noisy signal)が特定の多成分濃度推定値の質にどのように影響するかについての案内(guidance)をまったく与えない。

従って、多成分分析方法を用いて、スペクトル的に区別可能な複数の種を含む混合物のスペクトル応答に基づいて濃度推定値の質を決定する際の統計的に有意な効果尺度が必要とされている。

要旨

本発明は、混合物中に存在するスペクトル的に分解可能な複数の種のそれぞれの相対濃度の推定値の重複決定測定を用いる多成分分析処理を含む処理装置における信号処理方法の開示に関し、推定値の質を評価するための有効な効果尺度の決定を含む。

第1の局面において、本発明は、スペクトル的に分解可能な複数の分子種を含む混合物のスペクトル応答から生じる信号の多成分分析を行うための処理装置における信号処理方法を含む。この方法は、混合物中のそれぞれの種についての濃度推定値および統計的信頼区間を決定する。この方法は、以下のステップを有する。第1に、マルチチャネル検出器からデータベクトル d を受信する。但し、データベクトル d は長さ n_c を有し、 n_c はモニタしている検出器チャネルの数である。較正行列 K が与えられる。但し、 K は、 n_c 行および n_p 列を有し、 n_p は、スペクトル的に分解可能な分子種の数である。本発明の重要な特徴として、 n_c は n_p よりも大きく、従って、 K は非正方行列である。次に、濃度推定値ベクトル c を決定する。但し、 c は長さ n_p を有する。最後に、式

$$C I_i = c_i \pm (\text{var covar } (c_{ii}))^{1/2} Q(c_1, n_f)$$

に従って、濃度推定値ベクトルの各要素について信頼区間 $C I_i$ を決定する。但し、 Q は、所与の信頼水準 c_1 および所与の数の自由度 n_f についての統計分布の臨界値 (critical value) である。但し、 $n_f = n_c - n_p$ である。

本発明の好適な実施形態の1つにおいては、 n_c は10よりも大きく、より好ましくは、 n_c は20よりも大きい。

本発明の方法の第2の好適な実施形態においては、比 n_c / n_p を1.5よりも大きくなるように選択する。より好ましくは、この比は3よりも大きい。さらに好ましくは、この比を5より大きくなるように選択する。

好ましくは、この方法において使用される統計分布は、ステューデント t 分布またはガウス分布である。

時間に関連した複数のデータベクトルを処理する別の好適な実施形態の1つにおいては、定められた期間の間この方法を一定間隔で反復する。

濃度推定値ベクトルを決定するステップは、好ましくは、式、

$$\mathbf{c} = (\mathbf{K}^\top \mathbf{K})^{-1} \mathbf{K}^\top \mathbf{d}$$

を用いて行われる。

信頼区間を決定するステップは、好ましくは、

濃度推定値ベクトル \mathbf{c} の残差ベクトル \mathbf{r} を決定するステップであって、

$$\mathbf{r} = \mathbf{d} - \mathbf{K} \mathbf{c}$$

であるステップと、

データベクトル \mathbf{d} の分散 $\text{var}(\mathbf{d})$ を決定するステップであって、

$$\text{var}(\mathbf{d}) = \frac{\mathbf{r}^\top \mathbf{r}}{(n_s - n_v)}.$$

であるステップと、

濃度推定値ベクトル \mathbf{c} の分散－共分散行列 $\text{varcovar}(\mathbf{c})$ を決定するステップであって、

$$\text{varcovar}(\mathbf{c}) = \text{var}(\mathbf{d}) (\mathbf{K}^\top \mathbf{K})^{-1}$$

であるステップと、

を含む。

第2の局面において、本発明は、上記の多成分分析方法の方法ステップを実行するために機械によって実行可能な複数の命令からなるプログラムを実体的に具現化する、機械によって可読なプログラム格納装置を含む。

第3の局面では、本発明は、ポリヌクレオチド配列決定方法を含む。この方法では、標識されたポリヌクレオチドの第1、第2、第3および第4のクラスを含む混合物が形成され、第1のクラスのポリヌクレオチドの各々は、3'－末端ジデオキシアデノシンを含み、第1の標識で標識され、第2のクラスのポリヌクレオチドの各々は、3'－末端ジデオキシシチジンを含み、第2の標識で標識され、第3のクラスのポリヌクレオチドの各々は、3'－末端ジデオキシグアノシンを含み、第3の標識で標識され、第4のクラスのポリヌクレオチドの各々は、3'－末端ジデオキシチミジンを含み、第4の標識で標識される。これらの標識は、第1、第2、第3および第4の標識の各々がスペクトルで分解され得るように選

択される。標識されたポリヌクレオチドは、電気泳動により分離され、同様の大きさのポリヌクレオチドのバンドを形成する。バンドの各々は、 n_c 個のチャネルを有するスペクトルアレイ検出器を用いて検出される。ここで、 n_c は4よりも大きい。スペクトルアレイ検出器からの信号は、本発明の方法に従って処理され、 c_i およびC I_iの値は、一連の時間値について報告される。

第4の局面では、本発明は、スペクトル的に区別可能な n_p 個の種を含む混合物を含むサンプルのスペクトル応答を測定するための装置を含む。サンプルのスペクトル応答を検出するためのスペクトルアレイ検出器が設けられ、この検出器は、 n_c 個のチャネルを有し、ここで、 n_c は n_p よりも大きく、検出器の出力は、特定の時間における特定のチャネルでの光強度を表す信号である。スペクトルアレイ検出器からの信号を受け取るための信号処理装置が設けられ、信号処理装置は、本発明の信号処理方法の方法ステップを行うように構成される。

図面の簡単な説明

図1は、4色DNA配列決定方法において用いられる色素の組の発光スペクトルを示す。

図2は、本発明の多成分分析方法の好適な実施形態を一般的に説明するフローチャートである。

図3は、スペクトル的に区別可能な複数の蛍光種を含む混合物のスペクトル応答の測定のために有用なスペクトルアレイ検出システムの好適な光路を示す。

図4は、色素R6Gで標識されたA終端標準較正混合物の20チャネル電気泳動図を示す。

図5aおよび図5bは、従来の多成分分析方法(図5a)および本発明の改良された多成分分析方法(図5b)を用いたDNA配列決定実験の結果を比較する。

図6は、本発明の方法による、スペクトル的に区別可能な複数の分子種を含む混合物のスペクトル応答の測定および分析のために有用な装置を示す。

好適な実施形態の詳細な説明

以下、本発明の好適な実施形態を詳細に参照する。本発明の実施例は、添付の図面に示される。本発明を好適な実施形態について説明するが、これが、本発明

をこれらの実施形態に限定することを意図するものではないことが理解されるであろう。むしろ、本発明は、添付の請求の範囲で規定される発明の範囲内に含まれ得る改変、変形および同等物をカバーすることが意図される。

本発明は概して、多成分分析を行うための処理装置における信号処理方法に関し、この方法は、混合物内に存在する、スペクトル的に分離可能な複数の種の各々の相対濃度の推定と、そのような濃度推定値の各々についての統計的に有効な信頼区間の決定のために、優決定測定値 (overdetermined measurement) の使用を採用する。一般に、本発明の方法は、(i) 多チャネル検出器からのデータベクトル d を受け取るステップと、(ii) n_c 行および n_p 列を有する較正行列 K を与えるステップとを包含し、ここで、 n_c は n_p よりも大きく、 n_c はモニタされている検出器チャネルの数であり、 n_p はスペクトル的に分解可能な分子種の数であり、(iii) 長さ n_p を有する濃度推定値ベクトル c を決定するステップをさらに包含し、各行列要素 c の大きさは、特定の種 i の濃度推定値を表し、(iv) 以下の式に従って濃度推定値ベクトル C_I_i の要素の各々についての信頼区間を決定する工程をさらに包含し、

$$C_I_i = c_i \pm (\text{var covar } (c_{i:i}))^{1/2} Q(c_1, n_f)$$

ここで、 Q は、所与の信頼水準 c_1 および所与の自由度を示す数 n_f についての統計分布の臨界値であり、 $n_f = n_c - n_p$ である。図2は、本発明の多成分分析方法の1つの好適な実施形態のフロー図である。

I. 定義

特に言及しない限り、本明細書において用いられる以下の用語および表現は、以下の意味を持つことが意図される。

サンプルの「スペクトル応答」とは、サンプルが、波長を有する放射線を吸収、散乱および／もしくは放射する程度、ならびに／またはそのような吸収、散乱および／もしくは放射の大きさを指す。スペクトル応答を与えることができるプロセスの例としては、UV—可視光吸收度、化学ルミネッセンス、蛍光、電気化学ルミネッセンス、ラマン分光法などがある。

本明細書で用いられる「信頼区間」という用語は、特定の時間パーセンテージ

でのパラメータの真の平均値を含む値の範囲を指す。従って、95%の信頼水準、即ち、時間の95%でサンプルの平均純度が99.95%±0.25%であれば、同じ方法で推定される推定純度は、99.95%±0.25%の範囲であろう。

「多チャンネル検出器」という用語は、ダイオードアレイ、電荷結合素子(CCD)システム、光電子増倍管のアレイなどの、電磁放射に敏感な個々にアドレス指定可能な検出器要素のアレイを含む。

II. 多成分分析方法

A. 信号

本発明の信号処理装置は、スペクトル的に区別可能な複数の分子種を含む混合物のスペクトル応答をモニタする多チャネルスペクトルアレイ検出器によって生成される信号を受け取る。スペクトルアレイ検出器の出力の大きさは、光強度に比例する。測定値のチャネル位置は、直接またはスペクトル分離手段のいずれかの使用によって測定された光の波長に関連する。そのようなスペクトルアレイ検出器の一例は、Kargerら(Karger)により提供される。

B. データベクトル

データベクトル d は、長さ n_c を有するベクトルであり、ここで、 n_c は、多チャネル検出器によりモニタされている検出器チャネルの数に対応する。各ベクトル要素 d_i の大きさは、所与の時間の特定の検出器チャネルでの信号強度を表す。本発明の重要な特徴では、モニタされている検出器チャネルの数 n_c は、検出されている、スペクトル的に分解可能な種の数 n_p よりも多い。好ましくは、 n_c/n_p は1.5よりも大きく、より好ましくは、 n_c/n_p は3よりも大きく、最も好ましくは、 n_c/n_p は5よりも大きい。スペクトル応答のこの「オーバーサンプリング」は、濃度推定値ベクトルの残差ベクトルの計算を可能にする目的で、優決定線形系を構成するために必要である。 $n_c = n_p$ であれば、この系は、残差を推定することができない、正確に決定された系を表す。

C. 較正行列

較正行列 K は、 n_p 個の種の各々による、 n_c 個の検出器チャネルの各々で検出

された信号への個々の寄与を表す純粋成分の線形応答定数からなる $n_c \times n_p$ の非正方行列である。較正行列の要素の値 K_{ij} は、各検出器要素についてのそれぞれの種の純粋成分の応答を測定することによって決定される。好ましくは、Kの要素の値は規格化され、すべてのチャネルについての規格化された値の和が 1 にされる。そのような規格化は、異なるチャネルと異なる走査との間での結果の定量比較を容易にする。

D. 濃度推定値ベクトル

濃度推定値ベクトルは、スペクトルが重なる多数の種からの寄与を含むスペクトル応答に基づく、スペクトル的に分解可能な多数の種を含む混合物中の個々の種のそれぞれの濃度推定値を表す。データベクトルが d であり、較正行列が K あるとすると、1つの好適な実施形態では、 c は、以下の式を評価することによって確認される。

$$c = (K^T K)^{-1} K^T d$$

ここで、上付き文字「 T 」は、行列 K の転置行列を示す。濃度推定値ベクトル c は、長さ n_p のベクトルであり、以前と同様に、 n_p は、その濃度が推定される、スペクトル的に分解可能な種の数である。従って、濃度推定値ベクトルの要素の値 c_i は、 n_p 個の種の各々の濃度推定値を表す。

本明細書における説明および実施例では、濃度推定値ベクトル c および残差ベクトル r は、重み付けされない最小二乗重回帰 (Draper) の方法を用いて推定される。本発明の方法では、重み付けされる最小二乗回帰法 (Draper)、主成分 (principle component) 回帰法 (Jolliffe)、特性根／リッジ (latent root/ridge) 回帰法 (Draper)、ならびに主成分および／または因子分析から得られたスコアに基づく関係 (Sharaf) などの他の回帰方法を用いてもよいが、これらに限定されない。各々の場合、濃度推定値は、 r の推定値を与え、これは、その後、測定値の分散および c, σ_s の信頼区間を推定するために用いられる。重み付け

される最小二乗回帰は、典型的には、幾つかの測定値 (チャネル) がその他の測定値よりも高い分散を有する場合に用いられる。そのような場合、重み付け行列

は、等しくない分散を補償する (account for) ために用いられる。この重み付け行列は、その後、 c における分散と、 c, \bar{c} の信頼区間を推定するために用

いられる (Draper, p. 108-116)。主成分回帰は、親スペクトルが非常に相関している場合に用いられる。このアプローチでは、 d 、 K および c の間の関係は、因子／主成分スコア空間で表される (Jolliffe; Sharaf)。スコア空間から得られた回帰推定値は、その後、固有ベクトルを用いて c に変換される。

E. 各濃度推定値に対する信頼区間の決定

所与のデータベクトル d は、各検出器要素におけるサンプルのスペクトル応答の測定値を含み、較正行列 K は、各検出器要素におけるそれぞれの種の個々の純粹成分の応答を示し、濃度推定値ベクトル c は、成分種のそれぞれの推定濃度を示し、各濃度推定値に関連する信頼区間 C_I が決定される。

本発明による各濃度推定値 c_i に対して信頼区間 C_{Ii} を計算する第 1 のステップは、濃度推定値ベクトル c に関する残差ベクトル r を計算することである。一般に、残差ベクトルとは、パラメータの観察値を示すベクトルと、回帰方程式から計算されるパラメータの推定値を示すベクトルとの相違を示すベクトルのことである。従って、本方法において、残差ベクトルは、 $r = d - Kc$ となるような観察データベクトル d と、推定ベクトル Kc との間の相違である。

次に、上記で評価した残差ベクトル r の値に基づいて、分散 $\text{var}(d)$ がデータベクトルに対して計算される。ここで、 $\text{var}(d)$ は、特定の測定値、例えば、検出器に到達する迷光、検出器の回路における電気的ノイズ、および他の同様に制御されない実験ファクターに関連する機器および環境による不確定性の測定である。 $\text{Var}(d)$ は、以下の式によって与えられる。

$$\text{var}(d) = \frac{\mathbf{r}^T \mathbf{r}}{(n_c - n_p)}$$

ここで、 n_c および n_p は、上記のように定義される。正方較正行列を用いる方法 (即ち、 $n_c = n_p$) に対しては、残差は決定され得ない。なぜなら、上記の式における比が決定されていないからである。このように正確に決定される線形

系は、推定パラメータの質の「統計的に有効な」測定を提供しない。

次に、濃度推定値ベクトル c の分散—共分散行列が決定される (`varcovar(c)`)。ここで、`varcovar(c)` は、上記の測定誤差がどのように濃度推定値ベクトル c に反映されているかを示す尺度である。`Varcovar(c)` は、以下の式を用いて評価される。

$$\text{varcovar}(c) = \text{var}(d) (K^T K)^{-1}$$

最後に、濃度推定値ベクトル c の各要素 c_i に対する特定の信頼区間 C_{I_i} が、評価される。

$$C_{I_i} = c_i \pm (\text{varcovar}(c_{ii}))^{1/2} Q(c_1, n_f)$$

ここで、 Q は、所与の信頼水準 c_1 および所与の数の自由度 n_f に対する統計分布の臨界値であり、ここで、この場合において、 $n_f = (n_c - n_p)$ であり、`varcovar(c_{ii})` は、`varcovar(c)` 行列の i 番目の対角要素を指す。本明細書で用いる用語「臨界値」は、臨界値より大きい値に対する分布曲線下の積分面積がゼロ（1 - 信頼水準）と等しくなる統計分布の値を指す。例示的な統計分布としては、正規分布またはガウス分布、ポアソン分布、 χ^2 分布、F 分布、ステューデント t 分布、および二項分布が含まれる。選択される特定分布は、特定の応用に依存する。好ましくは、ガウス分布またはステューデント t 分布が使用される。より好ましくは、 $n_f < 30$ のとき、ステューデント t 分布が使用され、 $n_f > 30$ のとき、ガウス分布が使用される。例えば、ステューデント t 分布を用いる 16 の自由度を有する系に対する 99 % の信頼区間を決定するためには、 Q の値は 2.92 である。

I I I . 装置

1 つの局面において、本発明は、スペクトル的に区別可能な複数の分子種を含む混合物のスペクトル応答の測定および分析に有用な装置を含む。図 6 を参照すると、本発明の装置は、(i) 各チャネルに到達する光の波長に関連する複数の検出チャネルを有するスペクトルアレイ検出器 100 と、(ii) スペクトルアレイ検出器によって発生される信号に対する本発明の多成分分析処理を行うための信号処理装置 105 と、任意に (iii) スペクトル的に区別可能な複数の種

を部分的に空間分離する分離手段110とを有する。好ましい構成において、装置は、処理装置への送信前に、スペクトルアレイ検出器からの信号をデジタル化するデジタイザ115を有する。

A. スペクトルアレイ検出器

本明細書で用いる用語「スペクトルアレイ検出器」は、(i)回折格子、プリズム、光学フィルタと組み合わせたビームスプリッタなどのサンプルからのスペクトル応答を形成する光をスペクトル分離する手段と、(ii)多チャネル検出器と、(iii)必要に応じて、白熱バルブ、アーク灯、レーザ、レーザダイオードなどの光源と、(iv)光源からの光および／またはサンプルのスペクトル応答を規定するために用いられる光を方向づけ調整することが可能な関連の光学部品とを用いる検出器を指す。

図3は、実時間の蛍光に基づいた多色電気泳動スキャナにおいて有用なスペクトルアレイ検出システムの好ましい実施態様の光路の模式図である。検出システムは、蛍光励起光源として、例えば、488 nmおよび514 nmの波長で最大強度を有する40 mWの直径0.67 mmの偏光ビームを放射するアルゴンイオンレーザなどのレーザを用いる。レーザ(66)からの光は、レーザ光を所望の位置に方向づける調整可能に設けられた回転ミラー(68)から反射される。次に、望遠鏡レンズ(70)は、ビーム直径を約100 μmに減少させ、屈曲ミラー(72)は、光を直角に電気泳動媒体(104)に方向づける。レーザ励起蛍光ラベルから放射された光は、非球面集光レンズ(74)によって集光され、集光レンズは、検出器の方向に光をコリメートする。次に、放射光は、レーザ拒否フィルタ(76)を通って、屈曲ミラー(72)の周りを通過し、それによって、検出器に入る散乱レーザ光のレベルを減少させる。励起レーザ光は、非球面集光レンズ(74)の中心を透過するので、ある量のレーザ光は、検出器の方向にレンズ表面から直接反射され、不要なバックグラウンド信号が生成される。レーザ拒否フィルタ(76)の中心に設けられる屈曲ミラー(72)は、この反射光を集光路から偏向させ、検出器に入る反射光の量を減少させるように作用する。集光された放射光は、次に平凸レンズ(78)を通過する。平凸レンズは、放射光

を分光器(82)への入口に設けられているスリット(80)に焦点合わせする。 (分光器(82)は、 17 nm/mm の分散度を有する 405 g/mm 、 450 nm ブレーズ格子を用いる。) 分光器(82)を透過した後、光は、CCD(90)に当たる。放射光のスペクトル分離は、完全に行われることはまれであるため、各検出器チャネルに当たる光は、多数の種からの放射の組合せである。 CCD(90)からの出力信号は、信号処理装置(64)に送信され、次の多成分分析およびデータ提示が行われる。実時間に基づいて複数の電気泳動レンジを調べる(interrogate)ために、回転ミラー(68)およびコンピュータ(90)を除いた上記の光学システムは、電気泳動チャンバの幅にわたって走査され得る。

B. 信号処理装置

本発明の好ましい実施態様において、上記の多成分分析方法のステップは、本発明の多成分分析方法のステップを実施するために特に構成された信号処理装置105によって成し遂げられる。このような信号処理装置は、適切なソフトウェアによって駆動される一般的なマイクロプロセッサ、内蔵ファームウェアを用いる専用マイクロプロセッサ、または特定のデータ収取、アナログ-デジタル変換、行列操作、または方法によって要求されるフィルタリング動作専用カスタマイズされたデジタル信号処理回路(DSP)の形態を取り得る。

1つの好ましい実施態様において、信号処理装置は、(i) 信号および較正行列のデジタル表現し記憶するためのメモリ、および(ii) 方法の様々なステップを実行するための処理装置を有する。このような場合、上記の方法ステップは、機械によって可読なプログラム記憶装置において具現化される。このようなプログラム記憶装置は、コンピュータ可読媒体を有する。コンピュータ可読媒体には、磁気ディスクケット、磁気テープ、光ディスク、読み出し専用メモリ、直接アクセス記憶装置および他の同様な任意の媒体が含まれる。

C. 分離手段

本発明の好適な実施態様において、例えばクロマトグラフィーおよび電気泳動などの分離プロセスにより部分的に空間的分離された後に、スペクトル区別が可

能な種を検出する。特に好適な実施態様において、検出前に電気泳動により種を部分的に分解しておく。好ましくは電気泳動マトリクスのタイプは、約2～20重量パーセントの濃度（重量対体積）を有する、架橋または非架橋のポリアクリルアミドである。より好ましくは、ポリアクリルアミドの濃度は約4～8パーセントである。好ましくは特にDNA配列決定の場合において、電気泳動マトリクスは例えば尿素およびホルムアミドなどの鎖分離剤、あるいは変性剤を含む。そのようなマトリクスを構築するための詳細な手順は、ABI PRISM™ 377 DNA Sequencer User's Manual, Rev. A, January 1995, Chapter 2 (p/n 903433, The Perkin-Elmer Corporation, Foster City, CA(PE)) に記載されている。特定の分離において用いられる至適なポリマー濃度、pH、温度、変性剤濃度などは、分離される分子種のサイズ範囲、その組成、核酸の場合単鎖であるか2重鎖であるか、および電気泳動によって情報を得ようとしているクラスなどの、多くのファクタに依存する。従って、本発明のアプリケーションは、特定の分離について条件を最適化するために、標準的な予備試験を必要とし得る。例として、本発明に基づき、19部対1部のアクリルアミド対ビスアクリルアミドから作成された6パーセントのポリアクリルアミドをpH 8.3のトリスボレート(Tris-borate)EDTAバッファ中で調製したマトリクス中において、約20～300塩基の範囲のサイズを有するオリゴヌクレオチドを分離および検出した。

別の電気泳動篩マトリクスとしては、親水性ポリマーセグメントからなるクラスのコポリマーから形成されるミセルネットワークであって、均一なセグメント長を有し、複数の疎水性ポリマーセグメントが互い上に担持されかつ親水性ポリマーセグメントによって規則的な繰り返し間隔で空間を空けられたミセルネットワーク、例えば $(C_6F_{13})_2PEG35000$ 、 $(C_8F_{17})_2PEG35000$ など (Menchen 1994) が含まれる。

I V. バイオ分析アプリケーション

本発明の方法は、空間的に重畠したスペクトル的に分解可能な複数の分析物の同時検出を必要とする任意の方法によく適している。しかし本発明の方法は特に、電気泳動またはクロマトグラフィーなどの生化学的分離手順を経た後のスペクト

ル的に分解可能なクラスのポリヌクレオチドの識別において、同様なサイズ、配置、帶電、疎水性などの物性を有するターゲット物質の一連のバンドまたはスポットが、直線状または平面状に存在する場合に、よく適している。本明細書において、「バンド」の用語は、例えば電気泳動による色素一ポリヌクレオチド共役物の分離におけるような、同様または同一な物性に基づいて、分析物を任意に空間的グループ化または集合化したものを包含する。

様々な状況において、色々なクラスのポリヌクレオチドが現れ得る。本明細書において「フラグメント分析」法と呼ぶ好適なカテゴリーの方法において、標識されたポリヌクレオチドフラグメントを、標識プライマーまたは標識ヌクレオチドを用いたテンプレート-ディレクテッド(template-directed)酵素合成、例えば連結またはポリメラーゼ-ディレクテッド(polymerase-directed)プライマース伸長などによって生成する。フラグメントを電気泳動またはクロマトグラフィーなどのサイズ依存型分離プロセスに供し、複数のクラスのポリヌクレオチドを同時に分離し、異なるクラスをスペクトル的に分解可能な標識によって区別する。次いで分離後に、分離されたフラグメントを例えばレーザ誘起による蛍光により検出する。

そのようなフラグメント分析法の1つは、可変数のタンデムリピートあるいはVNTR (Weber, Caskey) に基づくものである。VNTRは、2重鎖DNAのうちの、特定の配列の複数のコピーを隣接して(adjacent)含む領域であり、繰り返し単位の数は可変である。これらの繰り返し領域の多くは豊富(abundant)に見られ、高度に多形性であり、ヒトゲノム中に均一に分布しているため、有用な遺伝子マーカーとして機能し得る。VNTR座位の例として、pYNZ22、pMCT118、およびApo Bがある。VNTR法の部分集合として、ミクロサテライトリピートまたは短タンデムリピート(STR)、すなわち短い(2~4塩基)の繰り返し配列を特徴とするDNAのタンデムリピートの検出に基づくものがある。ヒトにおいて最も豊富に散在している繰り返しDNAファミリーは、(dC-dA)_n--(dG-dT)_nジヌクレオチド繰り返しファミリーである((CA)_nジヌクレオチド繰り替えしファミリーとも呼ぶ)。これらは、ヒトゲノムにおいて50,000~100,000もの(CA)_n繰り返し領域と考えられ、典型的にはブロック当たりにつき15~30リピートと考えられている。好ましくは、VN

TR

またはSTR法において、色素標識PCRプライマーを用いて標識をポリヌクレオチドフラグメント中に導入する。VNTR法を行うための好適な方法において、単一の電気泳動レーンにおいて複数の座位を分離し、各座位に由来するフラグメントを異なるスペクトル的に分解可能な標識で標識する。複数の座位を单一のレーンで泳動させることにより、サンプルのスループットが大きく増加し、異なる標識を有する内部レーン標準を用い得る。

特に好適なDNA配列決定方法において、本発明により識別されるクラスは、末端ヌクレオチドにより定義される。すなわち、4つの可能な末端塩基とスペクトル的に分解可能な色素の組の要素との間に対応関係を確立する (Smith)。DNA配列決定において有用であるそのようなスペクトル的に分解可能な色素の組は、DNA配列決定分野において周知である (Smith, Menchen 1993, Lee)。より好ましくは、これらのクラスは化学的方法あるいは鎖終結(chain termination)方法によるDNA配列決定において現れ、最も好ましくは、これらのクラスは鎖終結法(すなわちジデオキシDNA配列決定またはサンガー配列決定)において現れる。本方法は、配列を決定したい一本鎖または2本鎖DNAテンプレートを用いた、インビトロにおけるDNAポリメラーゼによるDNA鎖の合成を包含する。合成は、オリゴヌクレオチドプライマーがテンプレートにアニールする単一の部位のみで開始される。合成反応は、DNA伸長の続行をサポートしないヌクレオチドアナログを導入することにより、停止される。鎖終結ヌクレオチドアナログ、3'から5' DNA鎖伸長に必要な3'-OH基を欠失した、2',3'-ジデオキシヌクレオシド5'-トリホスフェート(ddNTP)である。dNTP (2'-ジオキシヌクレオシド5'-トリホスフェート) および4つのddNTPのうちの1つが適切な割合で用いられたとき、酵素触媒されたポリマー化は、ddNTPを導入可能な各部位における鎖の全体集合のうちの1画分において終結する。各反応において標識プライマーまたは標識ddNTPを用いた場合、高解像度電気泳動による分離後に、蛍光により配列情報を検出することが可能である。鎖終結法において、本発明の色素は、配列決定プライマー (sequencing primer) またはジデオキシヌクレオチドに付着され得る。VNTR分析の場合

と同様に、1つ以上のクラスの標識ポリヌクレオチドを单一のレーンで泳動させることが特に好ましい。なぜなら、サンプルスループットが大きく増大し、レ

ーン間の電気泳動度(electrophoretic mobility)の変動に起因する曖昧さが排除されるためである。

V. 実施例

本発明を、以下の実施例を考察することによりさらに明らかにする。実施例は、本発明を例示することのみを目的としており、その範囲を何ら限定するものではない。

実験

以下の実施例において用いられるデータは、ABI Model 310 Genetic Analyzer (The Perkin-Elmer Corporation, p/n 310-00-100/120) を用いて収集した。Model 310は、多色蛍光検出を用いたキャピラリー電気泳動型システムである。この検出器は、蛍光発光をスペクトル的に分離するための回折格子および、スペクトル的に分離された光の強度を波長の関数として測定するためのCCD多チャネル検出器を含んでいる。データを収集するために用いられたCCD検出器は、525nmから663nmの範囲の波長に対応して4～20の異なる収集チャネルがサンプリングされるように動作された。それぞれの場合において、各チャネルは7nmの波長範囲を有していた。

実施例において用いられた分離媒体は、6.6M尿素、0.66Mピロリジノン、100mM TAPSバッファ（水酸化ナトリウムによってpH8.0に調整）とともに処方された、5%フルオロカーボン(fluorocarbon)系ミセルポリマーマトリクス[2.5%(C₆F₁₃)₂PEG35000 + 2.5%(C₈F₁₇)₂PEG35000] (Menchen 1994) であった。キャピラリー電気泳動法のキャピラリーは、長さ47cmかつ内径75μmであり、フルオロハイドロカーボン (J&W Scientificから入手可能なDB-210コーティング) で内部コーティングされていた。キャピラリーの注入端から検出器窓までの分離距離は36cmであった。キャピラリーの温度を制御するための加熱プレートは42°Cに設定された。電気泳動運転電圧は160v/cmであった。各電解質貯蔵槽は、100 mM TAPSバッファ (pH 8.0) を含有していた。サンプルを25μlのTemplate Suppression Re

agent (PE p/n 401674) 中に再懸濁した。

実施例 1

較正行列を生成する

以下の実施例は、4色DNA配列決定実験において、本発明の多成分分析方法において用いられる較正行列Kの用意について述べる。

Applied Biosystems Taq Dye Deoxy® Terminator Standardsを用いて、同封された指示 (PE p/n 401071) に従って較正行列を用意した。各標準混合物は、サンガーモデルDNA配列決定伸長反応(extension reaction)から得られたフラグメントからなっており、各混合物のフラグメントは異なる塩基、A、G、CまたはTで終結していた。各標準混合物のフラグメントを異なる、スペクトル的に分解可能な蛍光色素で標識した。Aで終結したフラグメントはローダミン6G (R6G) で標識し、Gで終結したフラグメントはローダミン110 (R110) で標識し、Cで終結したフラグメントはローダミンX (ROX) で標識し、Tで終結したフラグメントはテトラメチルローダミン (TAMRA) で標識した。各標準混合物をModel 310システム上で上述の条件下で別々に電気泳動させた (run)。行列標準を40V/cmで30秒間、キャピラリーに注入した。

図4は、R6Gで標識した、Aで終結した標準混合物を示すデータのセグメントを示す。図は、走査番号1520～走査番号1600までの、チャネル1～20から収集したデータを示す。上記各走査番号は約1秒に対応する。チャネル1～20の各々からのデータを図に示す。

較正行列用の、ベースライン修正した規格化データを得るために、以下の手順を用いた。まず、局所的な信号極大に対応する走査番号、例えばこの場合では走査番号1570を同定した。次いで、摂動のないベースライン値に対応する走査番号、例えばこの場合では走査番号1581を同定した。その後、2つの選択された走査番号の各々において、20の検出器チャネルの各々からデータを収集した。次に、信号極大データのベースライン修正するために、修正した走査値を、信号極大データからベースラインデータを減算することにより演算した。最後に

、全20のチャネルからの修正した走査データの総計が1となるように、例えば
二

の場合では修正した各走査データが36, 142で除算されるように、修正した走査データを規格化した。上記の手順は、各標準混合物用の較正スペクトルの、より確実な推定を得るために、他のピーク／ベースラインデータ対に関しても反復され得る。規格化した修正走査値を用いて、較正行列を生成した。以下の表1は、チャネル1～20の各々について、信号極大データ（走査1570）、ベースラインデータ（走査1581）、修正した走査データ、および規格化した修正走査データを示す。

表1

走査1570および1581に基づいて20のチャネルにおいて収集された、
信号極大データ、ベースラインデータ、修正したデータ、および
規格化した修正走査データ

チャネル番号	走査 1570	走査 1581	修正した走査	規格化した 修正走査
1	8369	8178	191	0.00528471
2	9693	9046	647	0.01790161
3	11254	9744	1510	0.041779647
4	12554	9898	2656	0.073487909
5	13351	9847	3504	0.096950916
6	13442	9683	3759	0.104006419
7	13498	10026	3472	0.096065519
8	13656	10508	3148	0.08710088
9	15412	12606	2806	0.077638205
10	14710	12272	2438	0.067456145
11	12932	10827	2105	0.058242488
12	12451	10465	1986	0.05494992
13	12289	10487	1802	0.04985889
14	12442	10872	1570	0.043439765
15	12443	11082	1361	0.03765702
16	11679	10650	1029	0.028471031
17	10671	9849	822	0.022743622
18	10104	9406	698	0.019312711
19	9044	8597	447	0.012367882
20	7830	7639	191	0.00528471

他のG、CおよびTの標準混合物の各々に関して、上記の手順を反復した。得られた較正行列Kの値を以下に示す。

K -			
0.06997656	0.00560156	0.0006317	0.00106842
0.12375633	0.01742844	0.00135174	-0.000216
0.13982928	0.04116091	0.00123986	0.0001631
0.12371946	0.07320856	0.00467437	5.78E-05
0.09886611	0.09564174	0.01468073	0.00110168
0.07713955	0.10223842	0.03331307	0.00135638
0.06114513	0.09631226	0.06274018	0.00233467
0.05158994	0.08743771	0.09389819	0.00552211
0.04562928	0.07895128	0.11459118	0.01693018
0.03976041	0.06822545	0.11740752	0.04309424
0.03399378	0.06118095	0.1077492	0.08495856
0.02968821	0.05432655	0.09114524	0.12773016
0.0260446	0.04939954	0.0760703	0.15276644
0.02068053	0.04293259	0.06362222	0.14843742
0.01719018	0.03699554	0.05479123	0.12737328
0.0138741	0.03001199	0.05043105	0.10408945
0.0108418	0.02420318	0.04193956	0.07498112
0.00823091	0.01703111	0.03416537	0.05370087
0.00570478	0.01182548	0.0237908	0.03455896
0.00233907	0.00588673	0.01176649	0.01999116

実施例 2

単一データベクトルの多成分分析

以下の実施例は、本発明の多成分分析方法の、4色DNA配列決定実験から得られたデータの分析への適用について述べる。本実施例は、上記実施例1において導出された較正行列を用いる。実施例は、本発明の多成分分析方法を、DNA配列決定実行中に収集された单一データベクトルdに適用する。

A p p l i e d B i o s y s t e m s D y e T e r m i n a t o r C

y c l e S e q u e n c i n g R e a d y R e a c t i o n K i t を用いて同封された指示 (P E p / n 4 0 2 0 7 9) に従って、4色色素標識終結反応物を調製した。テンプレートは p - G E M® プラスミド DNA であり、前

方向プライマー伸長反応を - 2 1 M 1 3 プライマーでプライムした。プライマー

伸長工程に続いて、反応物をエタノールで沈澱させ、乾燥生成物を、 $20\mu\text{l}$ のテンプレート抑制試薬中に取り出して 95°C で2分間変性させた。電気泳動を上述のように行つた。

走査1685で収集されたデータベクトルdを以下に示す。表に示すデータには、ベースライン修正および単位領域への規格化が施されている。各チャネルに対するベースライン修正は、(i) そのチャネルに対する最小値に対して実行全体を走査すること、および(ii) データベクトルdの対応する要素から最小値を減算することによって行った。

d =
0.0030
0.0057
0.0090
0.0143
0.0231
0.0391
0.0617
0.0873
0.1048
0.1089
0.1020
0.0881
0.0752
0.0638

0.0556
0.0494
0.0407
0.0323
0.0228
0.0131

古典的な重回帰方法を用いて、濃度推定値ベクトルcを演算した。ここで、

$$c = (K^T K)^{-1} K^T d$$

上記式において、Kおよびdの値は上記に示している。cの値を以下に示す。

c =
0.0320
0.0768
0.8480
0.0427

次いで、濃度推定値ベクトル c の残差ベクトル r を決定した。ここで、

$$r = d - K c$$

上記式において、 d 、 K および c の値は上記に示している。残差ベクトル r の値を以下に示す。

r =
-0.0002
-0.0008
0.0003
0.0007
0.0001
0.0005
-0.0009
-0.0009
-0.0006
0.0010
0.0012
0.0002

-0.0004
-0.0005
0.0004
-0.0006
-0.0003
-0.0005
0.0000
0.0017

次いで、データベクトルの分散、 $\text{var}(d)$ を計算した。ここで、

$$\text{var}(d) = \frac{\mathbf{r}^T \mathbf{r}}{(n_s - n_s)}$$

$\text{var}(d)$ の値は $6.3270 * 10^{-7}$ であることが判明した。

その後、濃度推定値ベクトルの分散—共分散行列である $\text{varcovar}(c)$ を以下の関係に従って評価した。

$$\text{varcovar}(\mathbf{c}) = \text{var}(\mathbf{d})(\mathbf{K}^T \mathbf{K})^{-1}$$

分散—共分散行列の値である $\text{varcovar}(\mathbf{c})$ を以下に示す。

$\text{varcovar}(\mathbf{c}) * 10^4 =$			
0.2441	-0.3311	0.1533	-0.0162
-0.3311	0.7301	-0.4594	0.0695
0.1533	-0.4594	0.4905	-0.1588
-0.0162	0.0695	-0.1588	0.1340

最後に、濃度推定値ベクトルの各要素に対する信頼区間 $C I_i$ を以下のように決定した。

$$C I_i = c_i \pm (\text{varcovar}(c_{ii}))^{1/2} Q(c_i, n_f)$$

本実施例において、統計分布としてステュードントの t 分布を用いた。表にした値を参照にし、 $16(n_c - n_p)$ の自由度と 99 % の信頼水準に基づいて、 Q の値を決定した。得られた Q の値は、2.92 であった。各 c_i に対するプラスおよびマイナスの信頼区間を以下に示す。

$C I_1 = 0.0320 \pm 0.0144$
$C I_2 = 0.0768 \pm 0.0250$
$C I_3 = 0.8480 \pm 0.0204$
$C I_4 = 0.0427 \pm 0.0107$

実施例 3

本発明の多成分分析方法を用いた改良された塩基コーリング

以下の実施例は、本発明の多成分分析方法の、4色DNA配列決定実験から得られた多数データベクトル（各走査番号ごとに1つのデータベクトル）分析への適用について述べる。本実施例は、上記実施例1において導出された較正行列を用いる。

図5は、公知の配列T A A T C A T G Gを示すDNA配列決定実験の一部として収集された電気泳動図の一部を示す。配列決定テンプレートは、 ϕ GEM[®]プ

ラスミドDNAであり、実験条件は上述した通りであった。図5aは、従来のデータ分析方法を用いて得られた。従来のデータ分析方法において、それぞれ53

5 nm, 555 nm, 575 nmおよび605 nmで4つの検出器チャネルを用い、各チャネルのバンド幅は7 nmであった。図5 a の重要な特徴は、G信号からの干渉が、走査番号2255近傍におけるC塩基の塩基割り当てを曖昧にすることである。4チャネルシステムは正確に決定されたシステムであるため、塩基がCまたはGであるという可能性を比較するために推定される信頼区間は得られない。

図5 bは、図5 aと同一の電気泳動図を示すが、本発明の多元決定された多成分分析方法を用いてデータが分析された点が異なる。図5 bにおいて、4ではなく20のチャネルをモニタした。図5 bのプロットは、16の自由度を有するステュードントのt分布に基づくデータの99%の信頼区間の下限のみを示す。図に見られ得るように、99%の信頼値のみを提示することにより、図5 aからの

データに基づいた場合に自身の同定が不確定であった塩基が、明瞭にCと同定され得る。

本明細書で引用した全ての文献および特許出願は、個々の文献および特許出願が特定的に及び独立して参考のため援用されると示された場合と同一のレベルで、参考のため援用される。

上記にはいくつかの実施形態のみを詳細に述べたが、当業者であれば、その教示から逸脱することなく好適な実施形態において多くの改変が可能であることを明瞭に理解する。全ての改変は以下の請求の範囲に含まれることが意図される。

【図1】

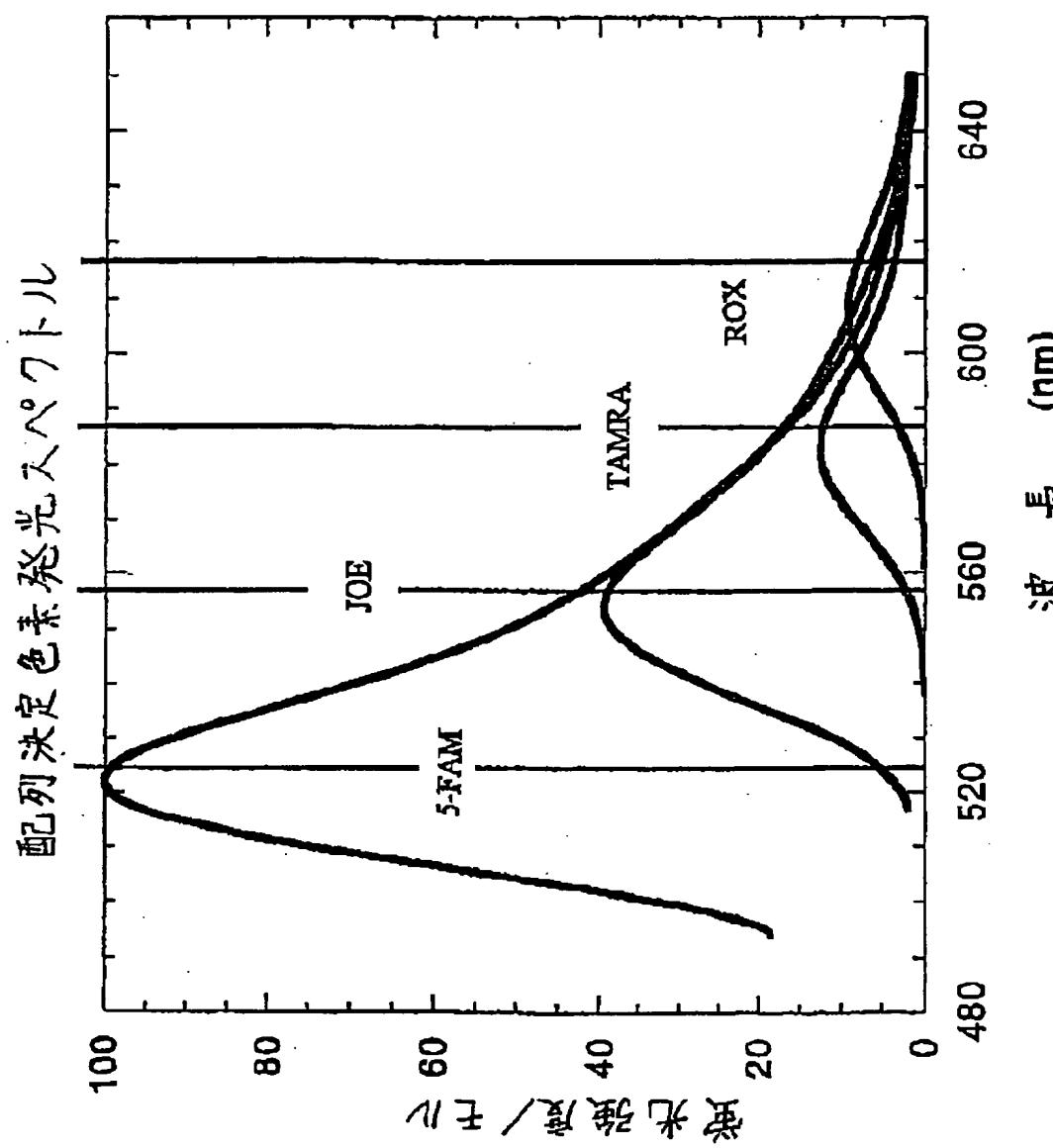


Fig. 1

【図2】

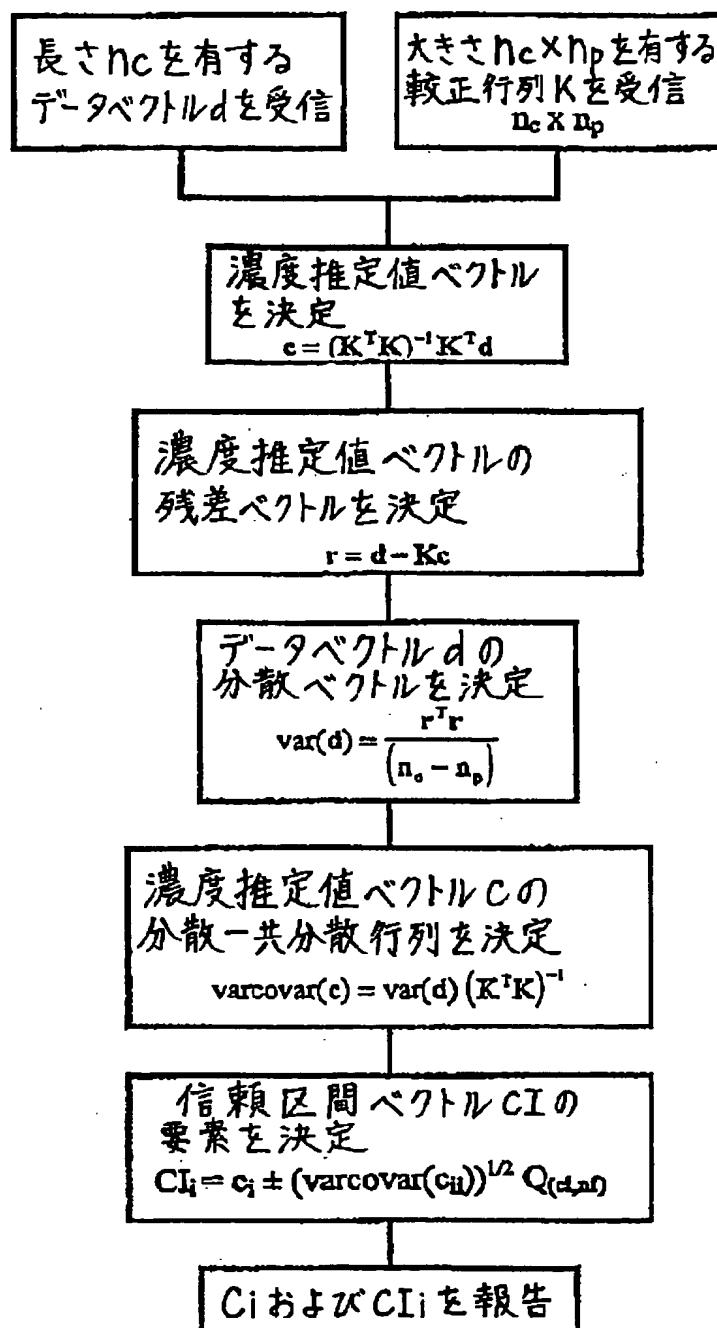


Fig. 2

【図4】

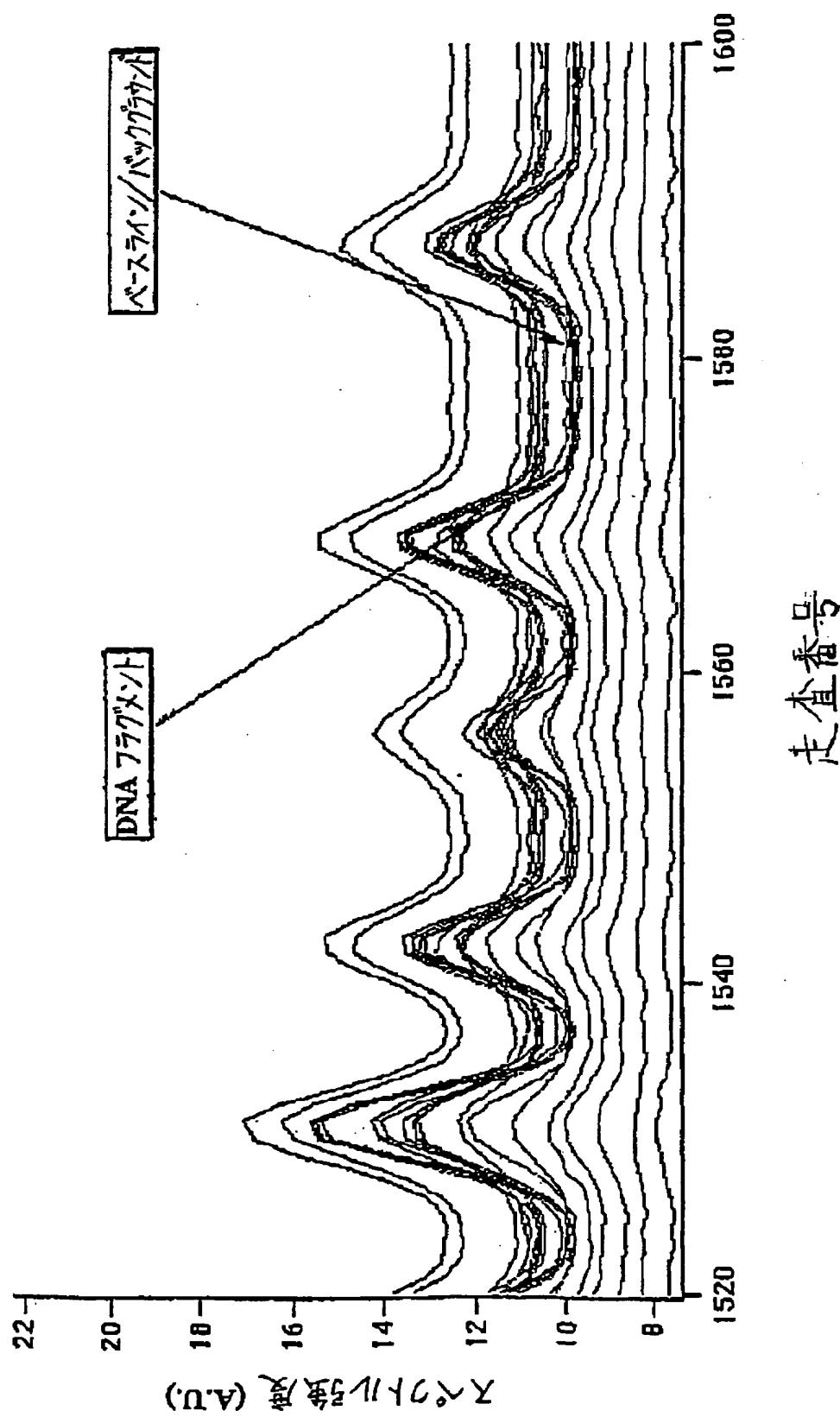


Fig. 4

【図5】

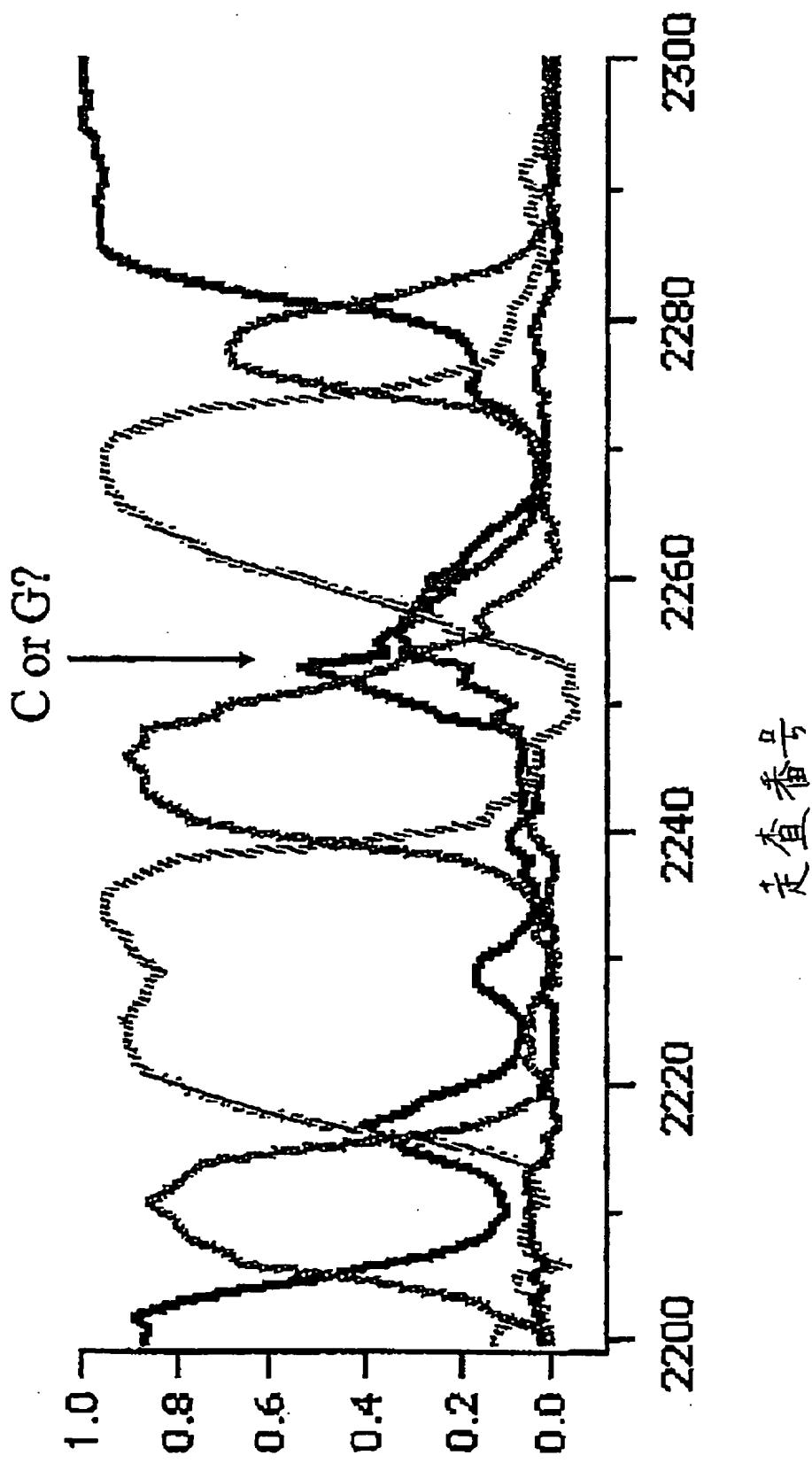


Fig. 5a

【図5】

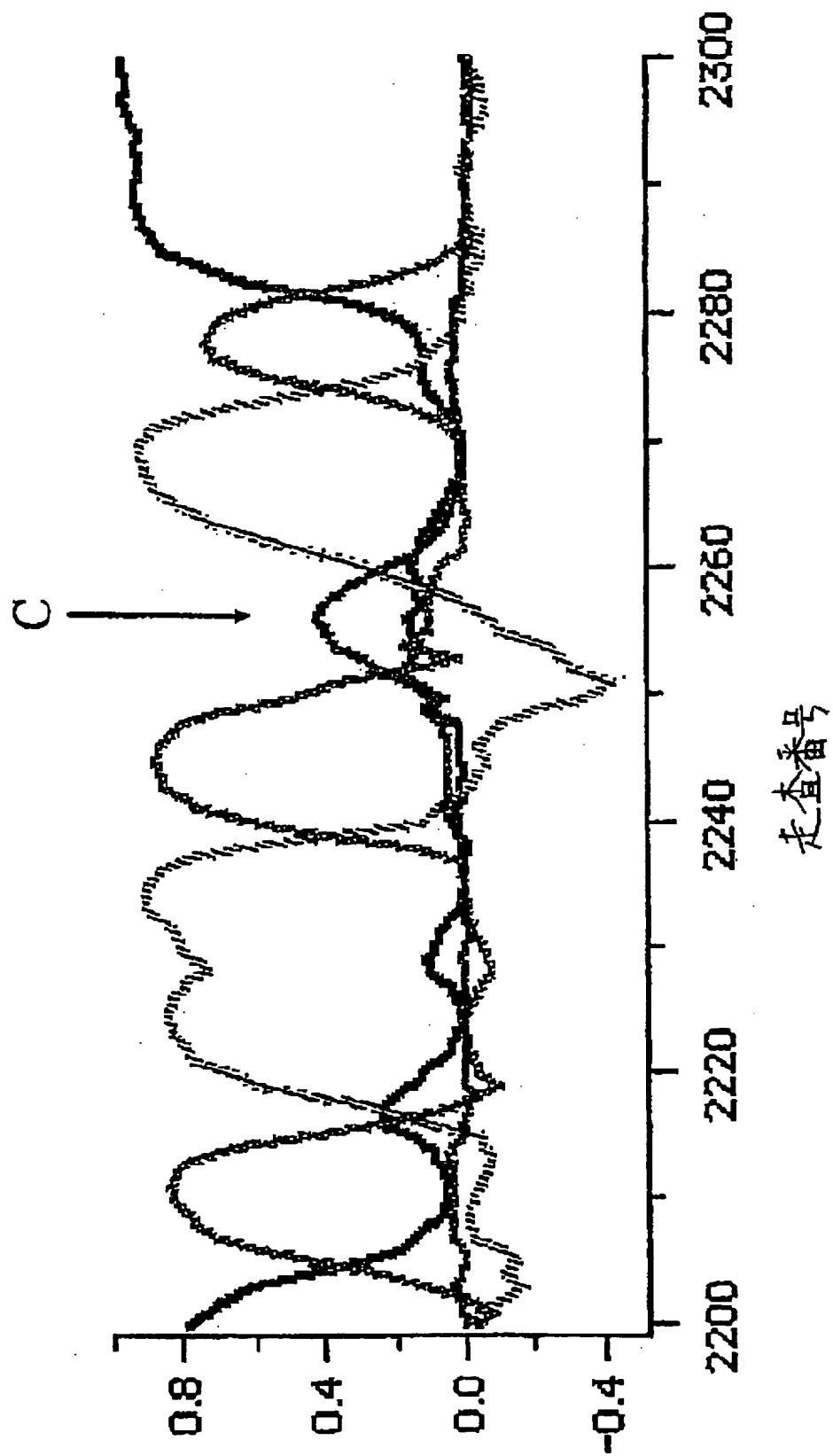


Fig. 5b

【図6】

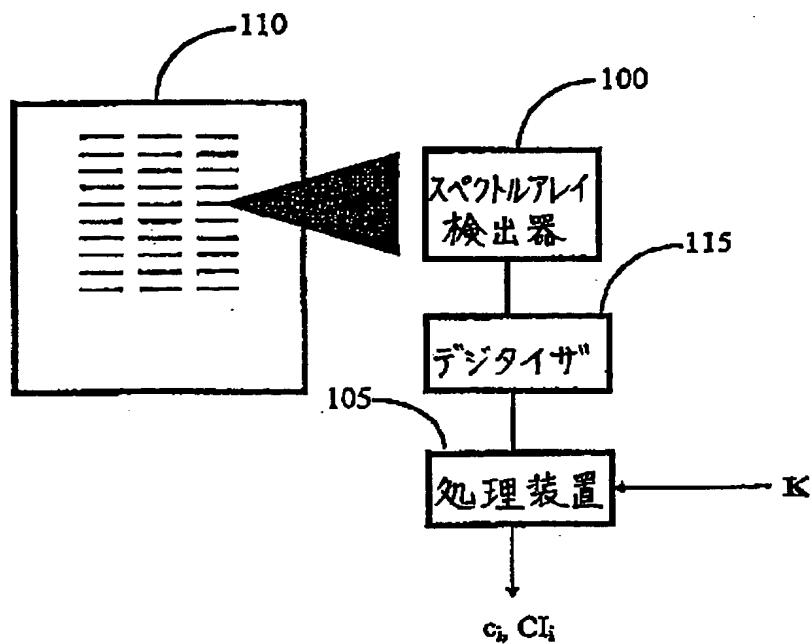


Fig. 6

【国際調査報告】

INTERNATIONAL SEARCH REPORT

Int'l. Appl. No.
PCT/US 97/08237

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G06F19/00 G01N30/86 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electroic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	OTTO ET AL: "selectivity in multicomponent analysis" ANALYTICA CHIMICA ACTA, vol. 180, 1986, AMSTERDAM THE NETHERLANDS, pages 445-456, XP002043255 cited in the application see the whole document ---	1,9, 11-13
A	SALAMIN ET AL: "a wavelength and optical path length selection procedure for spectroscopic multicomponent analysis" CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS, vol. 11, 1991, AMSTERDAM THE NETHERLANDS, pages 57-62, XP002043256 see the whole document ---	1,11-13 -/-

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "Z" document member of the same patent family

Date of the actual completion of the international search

10 October 1997

Date of mailing of the international search report

20.10.97

Name and mailing address of the IBA

European Patent Office, P.B. 581B Patentkantoor 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epofn,
Fax: (+31-70) 340-3016

Authorized officer

Guingale, A

INTERNATIONAL SEARCH REPORT

Inten	ntal Application No
PCT/US 97/08237	

C(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CONNELL ET AL: "automated dna sequence analysis" BIOTECHNIQUES. vol. 5, no. 4, May 1987, US, pages 342-348. XP002043257 cited in the application -----	12

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

フロントページの続き

(51) Int. Cl. ⁷	識別記号	F I	テ-マコード (参考)
G 0 1 N	27/447	G 0 1 N	33/50 P
	30/74		27/46 U
	30/86		27/26 3 2 5 A
	33/50		33/50 D
		C 1 2 N	15/00 A